

高遅延環境下における小粒度 iSCSI アクセスの特性解析

Analysis of iSCSI Storage Access with Short Blocks in Long-Latency Network

山口実靖¹
Saneyasu Yamaguchi

小口正人²
Masato Oguchi

喜連川優¹
Marusa Kitsuregawa

東京大学生産技術研究所¹
Institute of Industrial Science, The University of Tokyo

お社の水女子大学²
Ochanomizu University

1 はじめに

接続距離、管理の容易さなどの利点を持つ SAN として IP-SAN や iSCSI が注目を集めている。著者らは文献 [1] において高遅延環境下における iSCSI シーケンシャルアクセスのスループットの向上手法について述べた。本稿ではさらに、ショートブロックを用いた iSCSI アクセスのターンアラウンドタイム性能について考察する。

2 iSCSI 基本性能測定

以下の、3 実装を用いて小粒度 iSCSI アクセスの性能を測定した。① ニューハンプシャー大学 IOL が配布する draft 18 準拠の iSCSI 実装 (以下 “UNH 18” と呼ぶ)、② IOL 配布で draft 20 準拠の実装 (“UNH 20”), ③ Intel 社配布で draft 16 準拠の実装 (“Intel 16”). 簡略化のため iSCSI Target はメモリモードで動作させた。同環境における 8KB Read のターンアラウンドタイムを測定し図 1 の “UNH 18(def)”, “UNH 20(def)”, “Intel 16(def)” を得た。同図横軸は人工的に作成した Initiator-Target 間片道遅延時間である。同図よりターンアラウンドタイムは実装により 2 倍程度の差 (約 1RTT(Round Trip Time) と約 2RTT) があることが確認された。次に、これらの計測に対し開発した iSCSI 解析システム [1] を用いてそのパケットの転送を可視化した (図 2 参照) し “UNH 18” においては、Read 1 回につきネットワークの 2 往復が行われており、iSCSI Response が TCP Ack の受信を待って送信されていることが確認された。これは同実装において、TCP の Nagle のアルゴリズムが有効となっており、微小パケット iSCSI Response の送信が停止していることが原因である (“Intel 16” では TCP_NODELAY により Nagle のアルゴリズムが無効されており、“UNH 20” では Read 毎に Response を送信していない)。

3 TCP の振る舞いと HDD デバイスの影響の測定

次に、上記の実装に対し以下の改変を行いその性能を測定した。① “UNH 18” 実装に対し Nagle のアルゴリズムの無効化を行った (“UNH 18(ND)” と呼ぶ)、② iSCSI Data-in と iSCSI Response を結合し TCP 実装に対し 1 要求で受け渡し微小パケットが 2 個以上発生するのを回避した (“UNH 18(conc)”), ③ “Intel 16” 実装から TCP_NODELAY オプションを削除した (“Intel 16(N-ND)”). 同様に、そのターンアラウンドタイムを図 1 に記す。同図より、Nagle のアルゴリズム有効時における微小パケットの生成がターンアラウンドタイムの増加を招いていることが確認された。次に、実 HDD デバイスを用い

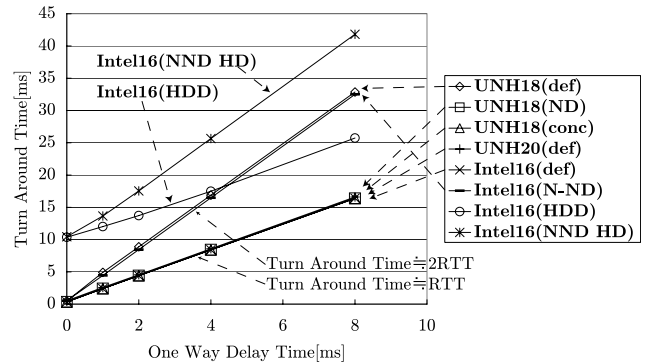


図 1 iSCSI Read (8KB) のターンアラウンドタイム

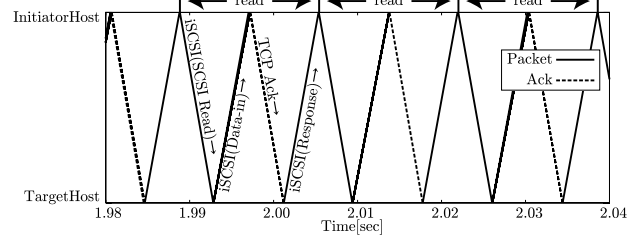


図 2 “UNH 18” におけるパケット転送の可視化

た計測を行い図 1 の “Intel 16(HDD)” (実 HDD を用いた “Intel 16” 実装), “Intel 16(NDD-HD)” (“Intel 16” から TCP_NODELAY を削除し実 HDD を使用) を得た。同測定はランダムアクセスであり Read 1 回毎に HDD アクセスを行っている。同図より実 HDD への応用時のターンアラウンドタイムは遅延時間 (1RTT または 2RTT) に固定的な HDD の応答時間を追加したものとなること分かる。一般に HDD の応答時間が 10ms 程度であるため遅延時間が 4ms を越える場合は本稿で述べたような往復回数の削減等が重要になると考える。

4 まとめと今後の課題

以上の様に、高遅延環境における小粒度 iSCSI アクセスのターンアラウンドタイムの縮小には、TCP/IP の振る舞いを考慮した iSCSI ドライバの実装が重要であり、その性能に大きな影響を与えることが確認された。今後は複数の iSCSI 接続を確立し、CPU 消費や TCP のハードウェア処理を考慮して単位時間あたりのトランザクション性能の向上について研究を進めていく。

参考文献

- [1] 山口実靖, 小口正人, 喜連川優. “iSCSI 解析システムの構築と高遅延環境におけるシーケンシャルアクセスの性能向上に関する考察”. 電子情報通信学会論文誌 D-1, 87, February 2004.